• NEWSLINE SPECIAL INSERT •

Up Close on genomics

A monthly insert on special topics at Lawrence Livermore National Laboratory. This month: The Human Genome Program. • • • April 2003

Focus On DNA & Genomics – Eddy Rubin

Offering profound insights into molecular functions

The completion of the Human Genome Project (HGP), announced earlier this week by the International Human Genome Sequencing Consortium, is one of the most significant milestones in the history of biology. Coming 50 years to the month after the publication of the double-helix structure of DNA by James Watson and Francis Crick, the HGP marks the beginning of an era that promises profound insights into the molecular functioning of all forms of life. Our understanding of cellular processes, the impact of organisms on each other and on the earth's environment, as well as fields of biological investigation yet to be identified, will be directly influenced by the discoveries and technologies of genomics.

As significant as it is, however, determining the genomic sequence of an organism is only one step along the path to understanding how the organism is built and how its actions are governed. The genome of an organism has been described as a "parts list"; as such, it helps us determine the basic elements involved in nearly all biological processes. The next major task for genomics is to begin drafting an "operating manual" that will tell us how the parts work together in their develop-

FOCUS, See page 8

Getting 'Up Close' with science

Editor's note: This Up Close edition focuses on genomics, a branch of bioscience celebrating the 50th anniversary of the double helix and completion of the project to sequence the human genome. The Joint Genome Institute (JGI), which unites the genomic programs at LLNL, Lawrence Berkeley and Los Alamos played a key role in the Human Genome Project. Up Close is a series spotlighting Laboratory programs and people. ◆

Solving life's secrets

How did life begin on our planet, and how did complex plants and animals evolve from simple, one-celled organisms? What causes birth defects and diseases, and why are some people more susceptible to disease than others? Why are some bacteria and viruses harmless, even beneficial, while others can be deadly? Scientists have been trying to solve these puzzles for centuries — and now, thanks to recent breakthroughs in biological, chemical, and genetic research and technology, the pieces are finally starting to come together.

U.S. Department of Energy Human Genome Program http://www.ornl.gov/hgmis

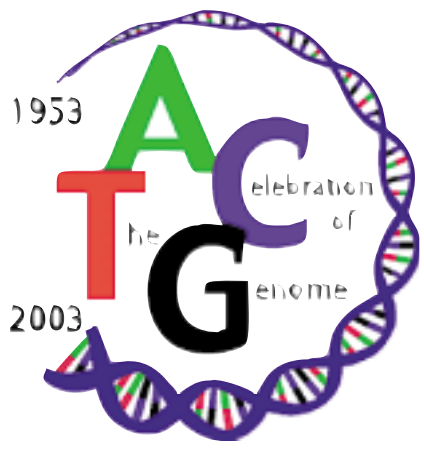
April 2003 marks two of the most important landmarks in this search for clues to the secrets of life: the completion this year of the human genome sequence, and James Watson's and Francis Crick's Nobel Prize-winning description of the DNA double helix 50 years ago.

On April 14, the U.S. Department of Energy (DOE) and the National Institutes of Health (NIH) announced the completion of the Human Genome Project — an international, $3-billion effort begun in 1990 to determine the complete sequence of the three billion DNA base pairs in the human genome. The news conference also described the federal government's vision for the future of genomics research and highlighted DOE's "Genomes to Life" program.

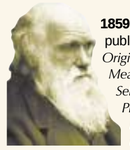
The DOE Joint Genome Institute (JGI) in Walnut Creek, a consortium founded in 1997 by Lawrence Livermore, Lawrence Berkeley and Los Alamos national laboratories contributed to the Human Genome Project by sequencing chromosomes 5, 16 and 19. Building on work begun at LLNL in the 1960s, the JGI has evolved in recent years from a single-purpose DNA sequencing facility to a full-fledged genomic research center. With the completion of the Human Genome Project, the JGI is now gearing up to sequence and study a wide variety of additional organisms whose genomes can shed light on the nature and functioning of the human genome, as well as on many natural processes that could provide insights into such DOE missions as energy production, environmental cleanup, and finding solutions to global climate change. ◆

The dawning of biotechnology

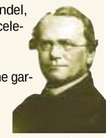
Perhaps the most important single clue to solving life's mysteries was made public 50 years ago this month, when James Watson and Francis Crick published a paper describing the now-famous double-helix structure of DNA - the "molecule of life." Building on that knowledge, scientists are learning more every day about how DNA works to determine the structure and functioning of all living things — and they are using that knowledge to prevent, diagnose and treat diseases, improve crop yields and the nutritional value of food, detect and clean up pollution, develop new energy sources, find clues to evolution, solve crimes, and even to fight terrorism. Much more is still to be learned, and some of the current and potential uses of this knowledge — such as cloning and genetic "screening" — are controversial and will be closely watched and debated as the research continues. But there is little doubt that the search for the secrets of life will continue and intensify during the 21st Century; "The Age of Biotechnology" is well under way. ◆



Cracking the Code



1859 Charles Darwin publishes *On The Origin of Species by Means of Natural Selection or The Preservation of Favoured Races in the Struggle for Life*. Although Darwin's landmark theory did not specify the means by which characteristics are inherited (because the mechanism of heredity had not been determined), his key premise was that evolution occurs through the selection of inherent and transmissible, rather than acquired, characteristics between individual members of a species.



1843–1868 Gregor Mendel, an Austrian monk now celebrated as "the father of genetics," conducts his experiments breeding the garden pea. Mendel established two laws that anticipated modern genetic research. The law of segregation states that the "factors" (what we now call genes) that determine such traits as height and eye color come in pairs, and the pairs separate when sperm and egg cells reproduce in the process called meiosis. As a result, each offspring gets one form, or allele, of the pair from each parent, which explains why children exhibit traits of both their parents. Mendel's law of independent assortment states that the pairs of alleles separate independently of each other during meiosis. "My scientific labors have brought me a great deal of satisfaction," Mendel wrote, "and I am convinced that before long the entire world will praise the result of these labors." Mendel's work, however, was largely ignored for 30 years.

1869 Swiss physician Frederick Miescher isolates DNA from human white blood cells and the sperm of trout; he calls the substance "nuclein."

1879 Walther Fleming, a German biologist, uses brightly colored dyes to help him observe long, thin threads in the nuclei of cells that appear to be dividing. These threads are later called chromosomes. In 1882, Fleming publishes a summary of the process, which he calls "mitosis."

1883 Francis Galton of England, a cousin of Charles Darwin's, coins the word and helps popularize the notion of eugenics. Eugenics, the theory of improving human "stock" through "selective breeding," ultimately leads to Nazi racial cleansing and forced sterilization laws in the United States, as well as modern prenatal testing and genetic counseling.

1900 Three European scientists, Hugo de Vries, Karl Correns and Erich von Tschermak, independently publish papers that confirm Mendel's Laws of Inheritance, giving Mendel's work its long-delayed recognition.

1902 Two cytologists, the American Walter Sutton and the German Theodor Boveri, reveal that genes are found on chromosomes, and that chromosomes come in pairs that are similar to each other.

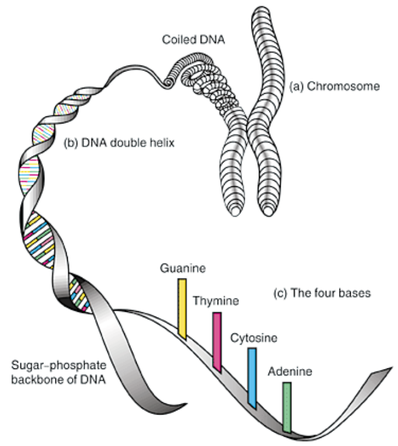
1910 Thomas Hunt Morgan and his team at Columbia University, and later CalTech, begin studying hereditary traits in Drosophila fruit flies. Their research reveals how genes are arranged in a row on chromosomes, as well as a variety of other genetic phenomena including sex-linked traits (traits that are passed only to one sex and not the other), the effect of a gene's location on its functioning, the existence of multiple alleles (gene forms), and chromosomal inversion (the reversal of a sequence of genes along part of a chromosome). Morgan's experiments also lead to *Drosophila's* unusual position as one of the best-studied organisms and most useful tools in genetic research to this day.

Genomics: a look at the sequences

Genomics is the study of the complete set of DNA sequences contained in each cell of an organism.

Found in every nucleus of every cell, the genome consists of tightly coiled threads of DNA (deoxyribonucleic acid) and their associated protein molecules, organized into structures called chromosomes. Except for mature red blood cells, all human cells contain a complete genome.

Genomics is different from genetics, which studies the inheritance and function of genes — segments of DNA that contain the chemical code that determines particular traits. Genes, however, make up only about three percent of human DNA; the rest is called "noncoding" DNA. Within these noncoding regions of the genome is the information that determines where and when genes are active, or "expressed." Genomics is the study of this much larger set of DNA sequences, both coding and noncoding.



DNA is made up of four similar chemicals (called bases and abbreviated A, T, C, and G) that are repeated millions or billions of times throughout a genome. The human genome, for example, has about three billion pairs of bases.

The genes in the DNA carry information for making all the proteins required by all organisms. These proteins determine, among other things, how the organism looks, how well its body digests food or fights infection, and sometimes even how it behaves. Proteins are large, complex molecules made up of smaller subunits called amino acids. Chemical properties that distinguish the 20 different amino acids cause the protein chains to fold up into specific three-dimensional structures that define their particular functions in the cell.

The particular order of As, Ts, Cs, and Gs along the DNA is extremely important. The order underlies all of life's diversity, even dictating whether an organism is human or another species such as yeast, microbe, fruit fly, or frog — all of which have their own genomes and are themselves the focus of genome projects. Because all organisms are related through similarities in DNA sequences, insights gained from nonhuman genomes often lead to new knowledge about human biology. ◆

Genome work is key to broader Laboratory missions

Genomics continues to play a key role in the Laboratory's rapidly evolving missions, notably in efforts to counter the threat of bioterrorism.

"Livermore has been involved in genome-related research since the 1960s. Our early work laid some of the foundations for the international effort to map the human genome, which was launched by DOE in 1986 and embraced by the the National Institutes of Health in 1988," said Bert Weinstein, acting associate director for Biology and Biotechnology Research Programs.

The Joint Genome Institute (JGI) was founded by the DOE in 1997 to consolidate the Human Genome sequencing efforts of Livermore, Los Alamos and Lawrence Berkeley laboratories. JGI played a key role in the project to map the human genome by sequencing Chromosomes 5, 16 and 19. Today the JGI can credibly claim to be the most efficient sequencing center in the world in terms of throughput and quality for the dollar.

The JGI works in complement with, and draws on, the resources of its founding labs. At Livermore this includes chromosome mapping, computational expertise and the ability to work with microbial pathogens and to provide finished sequence data starting with draft data produced by the JGI.

The ability to rapidly sequence the genomes of a variety of organisms is important to Livermore Lab missions in homeland security, as well as research in human health, bioremediation, climate change and energy, according to Weinstein.

"Good sequence data has become an essential and expected component of essentially all biological research. Genomics research is an integral part of Livermore lab programs in DNA assay development for Homeland Security, for understanding basic biological functions such as gene regulation, DNA repair, individual disease susceptibility, and protein structure and function determination," Weinstein said.

Looking to the future, genomic research will continue to be an important part of Livermore's biological research program. "The next phase in genomic research is learning how to extract more and more biologically relevant information from sequence data," he said, explaining that this effort includes comparative sequencing, particularly of regions of the mouse and other species, like chickens or fish, with a range of different evolutionary distances from humans.

"Having just the human sequence data is a little like having just one book in a language you don't understand," said Weinstein. "But if you have sequence data from many different species it begins to be like the Rosetta stone. The similarities and differences help us begin to understand not only the genes, but the regulatory elements that cause these genes to be turned on or off at particular times or in particular tissues in the body."

Microbial genomics is one of the newer LLNL bioscience thrust areas, and has promise for growth, Weinstein says. "Our focus continues on developing extremely sensitive and selective DNA-based identification methods integrated with rapid screening technologies for microorganisms that might be used in biological terrorism.

"Almost a decade ago we developed the technology for rapid amplification and identification of DNA using the polymerase chain reaction (PCR). Identifying a microbial pathogen can be done in as little as seven minutes following DNA extraction," Weinstein said.

"This technology has been the basis for a series of commercialized DNA detection technologies. In parallel, we have developed the ability to rapidly design assays that target specific microbes or viruses that are important either because of bioterrorism, human health or agricultural terrorism concerns. So we have an important role in both the 'hardware' and 'software' of DNA-based detection."

Developing these DNA assays depends on having good sequence data. As part of this effort, the Laboratory has completed sequencing *Yersinia pestis* (plague) and a very near relative, *Yersinia pseudotuberculosis*, and several other pathogenic and non-pathogenic bacteria.

"Throughout all of these developments, we have stressed the importance of outside validation and acceptance of our work," Weinstein said. "Our detection methods, as well as recently developed miniaturized detection devices, have been carefully evaluated at major field trials, and our DNA assays are validated in cooperation with the Centers for Disease Control and distributed to their nationwide Laboratory Response Network." ◆

Genomics: the essence of evolution

Since the first cell formed, the genome has been at the core of life's processes, and all living things are its products and its descendents. The pattern and processes of genetic inheritance and change are the essence of evolution.

Now that genome centers like JGI have developed the technology for very rapid DNA sequencing at remarkably low cost, evolutionary biologists have a chance to understand as never before the grand sweep of organic evolution — from bats to bacteria, from worms to wolves, from frogs to fruit flies.

Textbooks define the merger of evolutionary biology with genetics in the first half of the 20th Century as the "modern synthesis" period.

With centers like the Joint Genome Institute in Walnut Creek, above, the technology for rapid DNA sequencing can be performed at a low cost.

Now a new synthesis is beginning to appear at the boundary of evolutionary biology and genome science, with the potential for much greater understanding within each field. ◆

Sequencing grows by leaps and bounds

Biologist Susan Lucas, head of the Production Sequencing Department at the DOE Joint Genome Institute, has seen DNA sequencing technology come a long way in a big hurry.

When Lucas graduated from the University of Oregon and joined the bioscience staff at LLNL in 1997, new sequencing techniques and technologies were just being introduced that opened the door to automated, "high-throughput" sequencing – faster, cheaper, and more accurate – enabling an international team of researchers to complete the sequencing of the human genome this month, more than two years ahead of schedule.

Today, Lucas manages one of the world's fastest and most productive genome sequencing operations, sequencing more than 1.5 billion high-quality bases (Gb) of DNA – the equivalent of one-half of the human genome – every month.

The first methods for sequencing DNA were developed in the mid-1970s. At that time, scientists could sequence only a few base pairs a year, not nearly enough to sequence a single gene, much less the entire human genome. By the time the Human Genome Project began in 1990, only a few laboratories had managed to sequence a mere 100,000 bases, and the cost of sequencing remained very high.

Dr. Elbert Branscomb, an LLNL biomedical scientist who served as JGI's first director from 1996 to 2000, recalls meeting with eye-rolling skepticism from some of the world's leading genomics experts — including James Watson — when he predicted that JGI would produce 20 million "finished" bases in 1997 its first year of operation.

"We made our goal, hitting almost 21 million finished bases — this entailed the production of roughly 200 million raw bases — but we busted our butts to get there," Branscomb said. "Now look at what Susan and her team can do; they can sequence that many raw bases in less than three days —100 times faster and cheaper."

New high-speed capillary sequencing machines, improved robotics for DNA preparation and sequencing reactions, new purification methods, better project tracking, and incremental improvements in chemistry, all contributed to the increased efficiency and output of large-scale genomic DNA sequencing in this period. So did the development of whole-genome assembly software, which was critical to the success of the whole-genome shotgun sequencing strategy used in assembling a draft version of the human genome in 2000.

It wasn't just technology, however, that made it possible for the milestone achievement of the Human Genome Project to be reached so quickly. It also took careful planning, dedication, and concentrated effort by people like Lucas and her colleagues — spurred on by the desire to ensure that the "Book of Life" detailing the human genetic makeup would be freely available to researchers and the public.

"We worked really hard, because we wanted to see the public effort succeed," Lucas said. "The whole staff had a can-do attitude – we wanted to make it happen." ◆

Understanding microbes is key to fighting disease and pollution

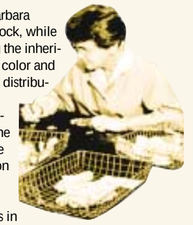
Thanks to its use of the latest and fastest sequencing technologies, the JGI is capable of draft sequencing the entire genomes of simple microorganisms such as bacteria, fungi and algae in a single day. This capability was put to the test when the JGI designated October 2000 "Microbial Month" and turned out high-quality draft sequences of 15 bacterial genomes in just 30 days.

Learning more about microbes is important not only to prevent and cure diseases that strike both humans and plants, but also to find ways to use microbes to break down industrial and hazardous wastes and other pollutants (bioremediation), absorb carbon dioxide and other gases that contribute to global warming (carbon sequestration), and enhance the productivity of organic energy and fiber sources (biomass and biofuels).

Cracking the Code

1926 Hermann Müller, a former member of Morgan's team, shows that exposure to X-rays can cause genetic mutations in *Drosophila*.

1941 While experimenting on bread mold, George Beadle and Edward Tatum show that genes regulate specific chemical events. They suggest that each gene directs the formation of one particular enzyme.



1944 Barbara McClintock, while studying the inheritance of color and pigment distribution in corn kernels at the Carnegie Institution Department of Genetics in Cold Spring Harbor, New York, discovers that genes can move from place to place on a chromosome and even jump from one chromosome to another. McClintock's discovery of transposable, or movable, genetic elements was greeted with initial skepticism but later recognized when, at age 81, she was awarded a 1983 Nobel Prize. Scientists now believe transposons may be linked to some genetic disorders such as hemophilia, leukemia and breast cancer. They also believe that transposons may have played critical roles in human evolution.

Oswald Avery, Colin MacLeod, and Maclyn McCarty of the Rockefeller Institute show that a molecule in the cell nucleus called deoxyribonucleic acid, or DNA — and not proteins, as previously believed — contains the factors that determine heredity in most organisms.

1952 Rosalind Franklin, a British chemist, uses a technique called X-ray diffraction to capture the first high-quality images of the DNA molecule.

1953 Franklin's colleague Maurice Wilkins shows the pictures to James Watson, an American zoologist, who has been working with Francis Crick, a British biophysicist, on the structure of the DNA molecule. After several false starts, Watson and Crick conclude that DNA is a double helix — two spiral strands that wind around each other like a twisted rope ladder.

1958 François Jacob and Jacques Monod predict the existence of messenger RNA, the molecule that carries information from the DNA in the cell's nucleus to the protein factories (the riobsomes) in the cytoplasm.

1962 Crick, Watson, and Wilkins share the Nobel Prize in medicine and physiology for the discovery that the DNA molecule has a double-helical structure. Rosalind Franklin, whose images of DNA helped

GENOMICS

The Human Genome and Beyond

What is genomics?

THE HUMAN BODY consists of trillions of cells. Almost all contain an entire **genome**–the complete set of inherited genetic information encoded in our DNA. When humans reproduce, the parents' sperm and egg DNA combine to contribute a genome's worth of genetic information to the fertilized embryo. That same information is in each of the cells that eventually make up an organism.

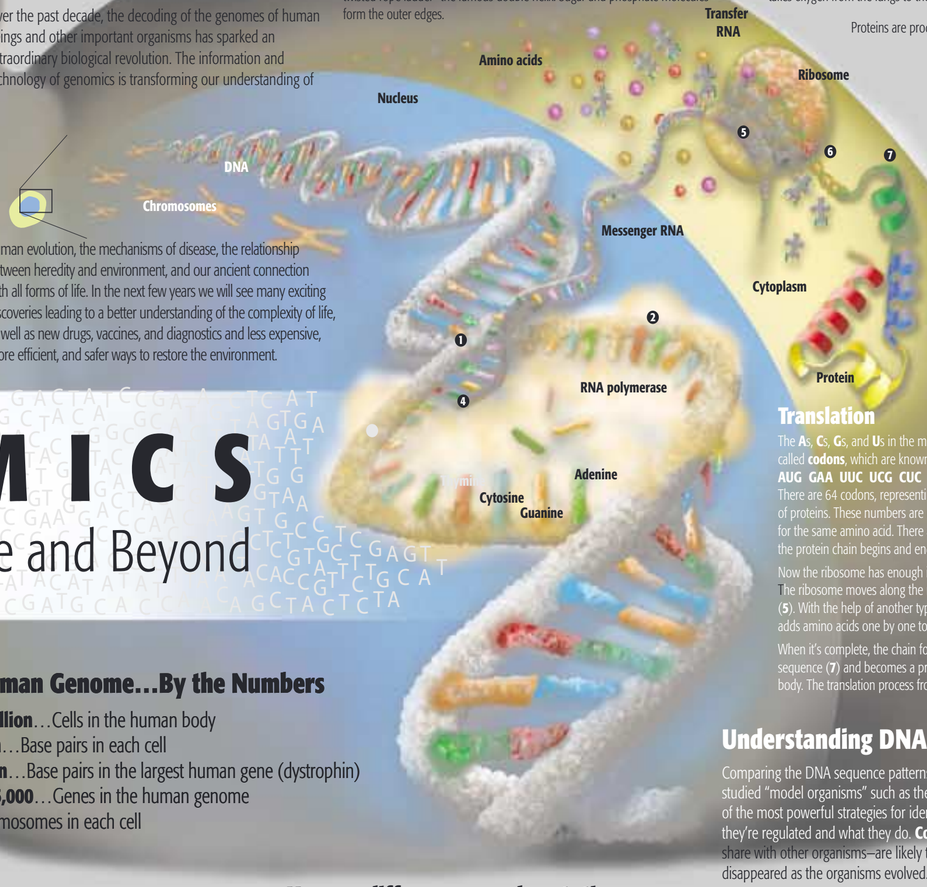
Some segments of DNA, called **genes** or "coding" DNA, contain the chemical recipe that determines particular traits; **genetics** is the study of the inheritance and function of these genes. Scientists now estimate that humans have about 30,000 genes, located along threadlike, tightly coiled strands of DNA called **chromosomes**. Genes, however, are only about three percent of human DNA; the rest is "noncoding" DNA. Within these noncoding regions of the genome is the information that determines when and where genes are active–for example, in which cell types and at what stages in the life of an organism. **Genomics** is the study of the entire set of DNA sequences–both coding and noncoding DNA.

Cell DNA Protein

Over the past decade, the decoding of the genomes of human beings and other important organisms has sparked an extraordinary biological revolution. The information and technology of genomics is transforming our understanding of human evolution, the mechanisms of disease, the relationship between heredity and environment, and our ancient connection with all forms of life. In the next few years we will see many exciting discoveries leading to a better understanding of the complexity of life, as well as new drugs, vaccines, and diagnostics and less expensive, more efficient, and safer ways to restore the environment.

DNA: life's code

The double-stranded DNA (deoxyribonucleic acid) molecule contains the four basic chemical units of life's code: the **nucleotide bases** adenine (**A**), guanine (**G**), cytosine (**C**), and thymine (**T**). These combine into the base pairs **AT**, **TA**, **GC**, and **CG**. The paired bases form the "rungs" of a structure that looks like a twisted rope ladder–the famous double helix. Sugar and phosphate molecules form the outer edges.



Translating the code: DNA ➡ RNA ➡ proteins

Following the DNA recipe, our cells manufacture the **proteins** that are responsible for the structure and functioning of our bodies. Proteins are involved in many of the body's life processes, including growth, repair, digestion, and aging. Many proteins are **enzymes** that can trigger or accelerate chemical reactions. Others are **transporters**, such as hemoglobin, found in red blood cells, which takes oxygen from the lungs to the body's cells.

Proteins are produced from the DNA recipe in two basic steps:

Transcription

Because DNA never leaves the cell's nucleus, a "messenger" must be created to move its information out into the cell. First, a key enzyme called **RNA polymerase** makes the DNA unwind and "unzip" by breaking the hydrogen bonds between the bases in the paired strands (**1**). This process forms two complementary strands–a **coding** strand (**2**) and a **template** strand (**3**).

RNA (ribonucleic acid), a single-stranded molecule very similar to DNA, is then created as nucleotide bases are strung together by the RNA polymerase in a sequence determined by the DNA template (**4**). The new RNA strand has the same information as the original coding strand, with one exception: **U** (uracil) nucleotides substitute for the **T**s in the DNA. The resulting strand is called **messenger RNA** (mRNA).

Next, the mRNA travels out into the body of the cell–the **cytoplasm**–and attaches to a **ribosome**, the cell's protein factory. Every cell has thousands of these tiny factories.

Translation

The **A**s, **C**s, **G**s, and **U**s in the mRNA are read by the ribosome as three-letter "words," called **codons**, which are known as the **genetic code**:

AUG GAA UUC UCG CUC

There are 64 codons, representing the 20 or so **amino acids** that are the building blocks of proteins. These numbers are not equal because sometimes two or more codons code for the same amino acid. There are also "start" and "stop" codons that determine where the protein chain begins and ends.

Now the ribosome has enough information to manufacture, or **synthesize**, a protein. The ribosome moves along the mRNA strand and reads its sequence, one codon at a time (**5**). With the help of another type of RNA called **transfer RNA** (tRNA), the ribosome adds amino acids one by one to the growing chain, called a **polypeptide chain** (**6**).

When it's complete, the chain folds into a specific shape dictated by the amino acid sequence (**7**) and becomes a protein; its shape determines the protein's function in the body. The translation process from DNA to protein is complete.

The web of life

We may not look alike, but humans, cows, fish, and microbes have a lot in common. We all retain similar DNA sequences inherited from our shared ancestors who lived hundreds of millions of years ago. Humans share many genes with mice, frogs, flies, and even bacteria and yeast, since their essential functions have been inherited intact over long periods of time.

Species	Chromosomes	Genes	Base Pairs
Human *(Homo sapiens)*	**46 (23 pairs)**	**28-35,000**	**~3.1 billion**
Mouse *(Mus musculus)*	**40**	**22.5-30,000**	**~2.7 billion**
Pufferfish *(Fugu rubripes)*	**44**	**~31,000**	**~365 million**
Malaria Mosquito *(Anopheles gambiae)*	**6**	**~14,000**	**~289 million**
Sea Squirt *(Ciona intestinalis)*	**28**	**~16,000**	**~160 million**
Fruit Fly *(Drosophila melanogaster)*	**8**	**~14,000**	**~137 million**
Roundworm *(C. elegans)*	**12**	**19,000**	**~97 million**
Bacterium *(E. coli)*	**1***	**~5,000**	**~4.1 million**

*Bacterial chromosomes are *chromonemes*, not true chromosomes .

The original estimate of more than 100,000 human genes was adjusted to between 28,000 and 35,000 when the draft human genome sequence was published in February 2001.

Genomics applications

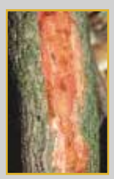
Fighting disease

Some human diseases and defects are directly or indirectly caused by genetic abnormalities. Sickle cell anemia, for example, is caused by a change in just one nucleotide out of six billion. Specific genes have been associated with breast cancer, deafness, and blindness. Some illnesses are caused by complex, interacting environmental and genetic factors and cannot be explained by classical inheritance patterns. Genome studies help medical researchers understand the molecular details of these diseases so they can pursue innovative drug treatments and more quickly identify high-risk individuals who could benefit from early medical intervention. And the analysis of the genomes of disease-causing microbes, viruses, and insects, such as the human malaria parasite and its carrier, the *Anopheles* mosquito, are helping in the development of new prevention and treatment strategies.

Most cystic fibrosis is from one DNA mutation–deletion of just 3 nucleotides–causing buildup of large amounts of mucus in the lungs.

Protecting plant life

Fungi and other plant pathogens cause billions of dollars in damage every year to agricultural crops, plants, and trees. Sequencing their genomes is helping botanists and foresters find effective treatments. Better understanding of plant genetics is also improving crop yields and enhancing the nutritional value of food.



An oak tree damaged by the sudden oak death pathogen Phytophthora ramorum

The Human Genome…By the Numbers

75-100 trillion…Cells in the human body
3.1 billion…Base pairs in each cell
2.4 million…Base pairs in the largest human gene (dystrophin)
28,000-35,000…Genes in the human genome
46…Chromosomes in each cell

Harnessing nature's technology

Microbes–nature's simplest and most abundant organisms–can thrive under extreme conditions of heat, cold, pressure, and even radiation. By studying their genomes, scientists hope to find ways to use bacteria and other microorganisms to solve a variety of environmental problems, develop new energy sources, and improve industrial processes. Some microbes can help clean up hazardous waste sites by absorbing, transforming, or breaking down contaminants–a technique called **bioremediation**. Others can help combat global warming by absorbing, or **sequestering**, carbon from the atmosphere. And microbes can convert a wide range of organic and inorganic materials into renewable energy.

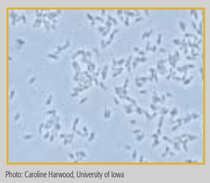
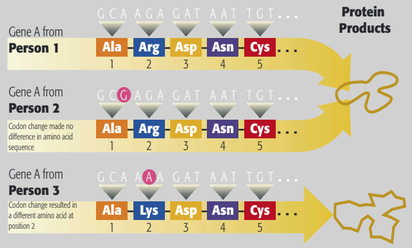


Photo: Caroline Harwood, University of Iowa

The bacterium Rhodopseudomonas palustris *can degrade complex aromatic hydrocarbons, assimilate carbon, and provide insights into biomass and biofuel production, particularly hydrogen.*

Human differences and mutations

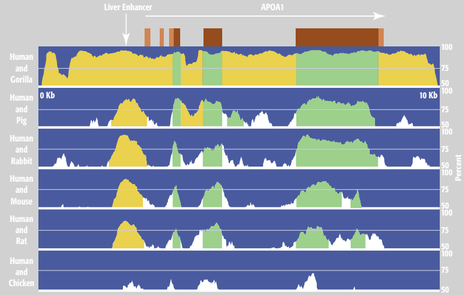
The DNA Sequence in every human is 99.9 percent identical to that of every other human. The slight variations in our genomes are called **single nucleotide polymorphisms**, or SNPs. Scientists estimate that there are about 1.4 million locations on the genome where SNPs occur in humans. It is these small variations that contribute to individual differences. SNPs and other mutations can be caused by copying errors as DNA is reproduced, or triggered by radiation, viruses, or toxic substances in the environment. Many SNPs have no effect on cell function, but others can cause or predispose a person to disease or influence response to a drug.



DNA sequence variation in a gene can change the protein produced by the genetic code.

Understanding DNA

Comparing the DNA sequence patterns of humans side-by-side with those of well-studied "model organisms" such as the fruit fly, mouse, pufferfish, and sea squirt is one of the most powerful strategies for identifying human genes and determining how they're regulated and what they do. **Conserved** sequences–DNA patterns that we share with other organisms–are likely to have important functions or they would have disappeared as the organisms evolved.

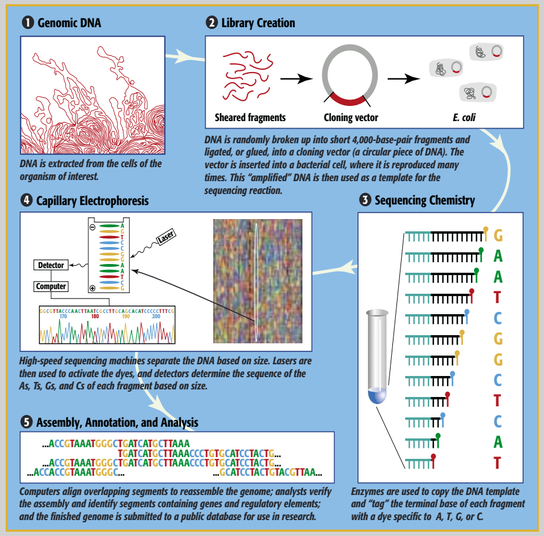


This computer plot shows how similar a segment of DNA is in the genomes of the human, gorilla, pig, rabbit, mouse, rat and chicken. Visualizations like these make it easier for scientists to identify conserved regions of DNA that could be important in regulating gene and protein function (green areas indicate conserved coding gene segments; yellow areas are conserved noncoding elements).

Along with helping identify genes and their functions, comparative genomics is shedding light on the functions of the noncoding sequences of DNA found within and between the genes. These segments can regulate gene **expression**, the process involved in determining when and where in the organism a given gene is turned on or off. Understanding the complex orchestration of gene and protein networks is a crucial aspect of contemporary biomedical research.

Sequencing genomes

While the number of chromosomes, genes, and base pairs in the genomes of different organisms vary, their fundamental structures are very similar, and the techniques for sequencing and studying them are the same. Whole genome shotgun sequencing is used to determine the order of the bases of an entire genome.



Field test version (March 2003). To comment e-mail: www@cuba.jgi-psf.org or call (925) 296-5808.

Prepared by the Department of Energy Joint Genome Institute and the Technical & Electronic Information Department, Lawrence Berkeley National Laboratory, with support from the Department of Energy Office of Science, Biological and Environmental Research Program. To order copies visit the poster website at www.jgi.doe.gov/education/posters.html.

Cracking the Code

lead to the discovery, died of cancer in 1958 and, under Nobel rules, was not eligible for the prize.

1966 Marshall Nirenberg and colleagues crack the genetic code by demonstrating that a specific sequence of three nucleotide bases (a codon, or nucleotide "word") codes for, or specifies, each of the 20-some amino acids used by the cell to produce proteins.

Robert William Holley, an American biochemist, shows that transfer RNA is involved in the assembly of amino acids into proteins. In the process, Holley becomes the first person to determine the complete sequence of a nucleic acid.

Hamilton Smith discovers the first restriction enzyme, a kind of molecular "scissors" that cuts DNA at specific points, in *Haemophilus influenzae* bacteria.

1972 Paul Berg and colleagues combine DNA from different species and insert it into a host cell, creating the first recombinant DNA molecules.

1973 Stanley Cohen and Herbert Boyer "cut and paste" DNA from a frog into an *E. coli* cell where it is reproduced, marking the dawn of genetic engineering.

1974-1975 Allan Maxam and Walter Gilbert of Harvard University develop the method of DNA sequencing bearing their name. The Maxam-Gilbert method uses phosphorous labeling and four separate chemical reactions to determine the sequence of DNA nucleotides. At the same time, Frederick Sanger and colleagues at the Laboratory of Molecular Biology, Cambridge, England, develop the "chain termination" method of DNA sequencing, which becomes, with slight modifications, the standard sequencing method used today.

1975 Concerned about possible risks from genetic engineering, 150 molecular biologists meet at the Asilomar Conference Center in Pacific Grove, California, to discuss ways to control gentic research until its hazards are better understood. Their recommendations lead to years of government supervision of recombinant DNA research until it is determined to be safe.

1976 Genentech, the first company devoted to genetic engineering, is founded in South San Francisco by Herbert Boyer and Robert Swanson.

1978 The gene for human insulin is cloned.

1980 Martin Cline and fellow scientists successfully transfer functional genes from one animal to another, creating the first transgenic mouse.

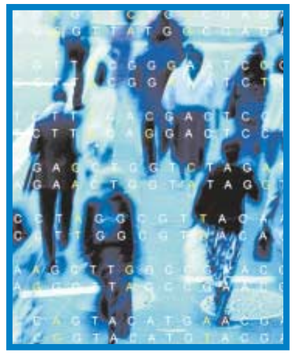
A human gene, coding for the protein interferon, is successfully introduced into and produced by bacteria.

After genetically engineering a bacterium capable of breaking down crude oil, Ananda Chakrabarty seeks to patent his creation under a provision of patent law providing patents for people who invent or discover any new and useful "manufacture" or "composition of matter." A patent examiner and the Patent Office Board of Appeals reject the patent on the grounds that living things are not patentable. The decision, however, is reversed by the U.S. Supreme Court in a 5 to 4 decision. The Court rules that while natural laws, physical phenomena, abstract ideas, or newly discovered minerals are not patentable, a live artificially-engineered

How we decode the human genome

The JGI's initial goal was to identify the genes and determine the sequence, or arrangement, of the DNA subunits (base pairs) in chromosomes 5, 16, and 19 — the Department of Energy's contribution to the international Human Genome Project, which DOE helped to launch in 1990 in order to learn more about how human health is affected by radiation, energy use and energy-production technologies. The three chromosomes contain 3,000 to 4,000 genes, including those whose defects may lead to genetically linked diseases such as certain forms of kidney disease and cancer, hypertension, diabetes and atherosclerosis.

As new technologies became available that greatly accelerated the sequencing process, JGI was able to complete the draft sequence of its three chromosomes in April 2000, several months ahead of schedule. Work to fill in the gaps in the draft sequence has continued into this year, and on April 14 the International Human Genome Project announced that the human genome sequence was 98 percent complete. The announcement came just two weeks before the 50th anniversary of James Watson's and Francis Crick's landmark paper describing the double-helix structure of DNA.



With the bulk of its work on chromosomes 5, 16, and 19 completed, JGI has turned its attention to decoding the genomes of a variety of microbes — work that could play an important role in fighting disease, pollution, global warming, crop damage and bioterrorism.

More recently, JGI has launched a research-driven program of comparative and functional genomics — studying the genomes of a number of "model organisms" in an effort to better understand the evolution, regulation and functioning of the genes and proteins that make life possible. ◆

Anticipated benefits of genome research

Rapid progress in genome science and a glimpse into its potential applications have spurred observers to predict that biology will be the foremost science of the 21st century. Technology and resources generated by the Human Genome Project and other genomics research are already having a major impact on research across the life sciences.

A look at some current and potential applications of genome research

Molecular medicine
- improved diagnosis of disease
- earlier detection of genetic predispositions to disease
- rational drug design
- gene therapy and control systems for drugs
- pharmacogenomics "custom drugs"

Microbial genomics
- new energy sources (biofuels)
- environmental monitoring to detect pollutants
- protection from biological and chemical warfare
- safe, efficient toxic waste cleanup
- understanding disease vulnerabilities and revealing drug targets

Risk assessment
- assess health damage and risks caused by radiation exposure, including low-dose exposures
- assess health damage and risks caused by exposure to mutagenic chemicals and cancer-causing toxins
- reduce the likelihood of heritable mutations

Bioarchaeology, anthropology, evolution, and human migration
- study evolution through germline mutations in lineages
- study migration of different population groups based on female genetic inheritance
- study mutations on the Y chromosome to trace lineage and migration of males
- compare breakpoints in the evolution of mutations with ages of populations and historical events

DNA forensics (identification)
- identify potential suspects whose DNA may match evidence left at crime scenes
- exonerate persons wrongly accused of crimes
- identify crime and catastrophe victims
- establish paternity and other family relationships
- identify endangered and protected species as an aid to wildlife officials (could be used for prosecuting poachers)
- detect bacteria and other organisms that may pollute air, water, soil, and food
- match organ donors with recipients in transplant programs
- determine pedigree for seed or livestock breeds
- authenticate consumables such as caviar and wine

Agriculture, livestock breeding, and bioprocessing
- disease-, insect-, and drought-resistant crops
- healthier, more productive, disease-resistant farm animals
- more nutritious produce
- biopesticides
- edible vaccines incorporated into food products
- new environmental cleanup uses for plants like tobacco

Source: DOE Human Genome Management Information System

Somewhere in our past, we're all related

Nature doesn't fix what isn't broken. That's why such different creatures as bacteria, fish, mice and humans have retained, or conserved, most of the same genes during our evolution from our common ancestors. No matter how different we look, all living organisms share most of the same basic biological functions; and how these functions operate is largely determined by the genes in our DNA.

Sometimes, however, something goes wrong — genes don't get copied correctly as organisms reproduce, or radiation or toxic substances in the environment trigger changes in otherwise healthy DNA. Microbes, viruses, and rogue molecules cause genes to get switched on or off at the wrong time. Mutations occur that can have drastic consequences: birth defects, genetic diseases, susceptibility to environmental factors that can result in other diseases, learning disabilities and behavioral or psychological problems.

Researchers can learn only so much about these genetic anomalies by studying the genome of a single organism. The Human Genome Project, for example, has zeroed in on the location of most of the 30,000 or so genes in the human genome — but scientists have only scratched the surface in determining what our genes do, how they evolved, how they work, how they repair themselves and what can happen if they fail to function properly. An even bigger mystery is the role played by the repetitive, non-coding so-called "junk" DNA that makes up about 97 percent of the human genome.

That's where comparative genomics comes in. By comparing the genomes of various organisms, linked by evolution, researchers can identify common regions and gain insights into how genes are regulated, or switched on and off, as well as how different genes affect the development and functioning of different kinds of organisms.

Finding genes that are unique to a particular species or strain can help medical researchers identify dangerous bacteria and viruses and develop techniques for preventing or treating the diseases they cause.

Comparative genomics has also enabled scientists to predict the existence of a substantial number of human genes the Human Genome Project has not yet tracked down. Thus, one of the Joint Genome Institute's key goals is to sequence the genomes of a variety of model organisms and make the information available to researchers around the world, so they can use it to help determine how DNA and genes work — and why they sometimes don't work the way they should. ◆

DNA clues from a deadly fish

JGI's sequencing and analysis of the pufferfish *Fugu rubripes*, a Japanese delicacy that can be poisonous if not prepared properly, made headlines around the world when it was published on the *Science* website in July 2002. Known as the "*Reader's Digest* version of the human genome," the Fugu genome contains most of the same genes and regulatory networks as the human genome, but in a DNA "package" about eight times smaller. The compact nature of the Fugu genome enabled researchers to predict the existence of almost 1,000 human genes that had not previously been identified. The analysis showed that nearly three-fourths of the genes in the human genome have identifiable counterparts in Fugu, highlighting the shared anatomy and physiology common to all vertebrates. The pufferfish project has laid the groundwork for researchers to begin identifying critical sections of DNA that control how genes direct where, when, and in what quantity proteins are manufactured. ◆

Surprising sea squirt

While it looks like a squishy blob, the humble sea squirt, *Ciona intestinalis*, is actually a surprisingly advanced organism whose tadpoles share a similar body structure with all vertebrates, including humans — they even have a primitive notochord, or backbone. Squirts also have an immune system and many of the same genes that humans do — including genes that can cause disease when they malfunction. Yet squirts have only a few thousand cells (compared with billions in complex vertebrates), making them an excellent "model organism" to study cell development and evolution and to learn how gene and protein regulatory networks operate — and if they might work the same way in humans. ◆

African frog hops into the picture

Building on their work with the sea squirt, JGI researchers are now tackling a more complex, human-like genome — that of the African clawed frog *Xenopus tropicalis*. By examining the same regulatory networks they have studied in the sea squirt, the scientists will learn how to "read" these networks like electricians read circuit diagrams. Frogs like *Xenopus* are popular with biologists because their growth from eggs to tadpoles to mature organisms sheds light on how cells divide and communicate with each other — the very processes that break down when birth defects occur or cancer strikes. By sequencing the *Xenopus* genome, which is about half the size of the human's, researchers hope to gain important new insights into how these processes work. ◆

Cracking the Code

microorganism is.

1982 The U.S. Food and Drug Administration approves the first recombinant DNA medical product, bacterially produced human insulin.

1983 Kary Mullis and colleagues of the Cetus Corp., Emeryville, California, invent polymerase chain reaction (PCR). This process, which allows the rapid reproduction of small samples of DNA, is applied within most facets of recombinant DNA technology, forensic analysis, and high-speed genome sequencing. Mullis received a 1993 Nobel Prize for his invention.

1984 A conference held by the U.S. Department of Energy (DOE) in Alta, Utah, discusses the possibility of using DNA research to detect tiny genetic mutations in the survivors of the Hiroshima and Nagasaki atomic bombs and their descendents. The conference sows the seeds for DOE's involvement in the Human Genome Project.

"Genetic fingerprinting," the technique of using sequences of DNA for identification, is developed by British geneticist Sir Alec Jeffreys. The entire sequence of the HIV-1 genome is determined by Chiron Corp.

1986 The Department of Energy announces its Human Genome Initiative, the genesis of the International Human Genome Project.

The first genetically engineered vaccine, a vaccine for hepatitis B, is approved by the Food and Drug Administration.

1987 Advanced Genetic Sciences conducts the first field trial of a recombinant organism (a bacterium) on an agricultural product (strawberries).

1988 The National Research Council endorses a national effort to sequence the human genome, and the National Institutes of Health establishes its Office of Human Genome Research, with James Watson as its first director.

1990 The Department of Energy and the National Institutes of Health formally launch the Human Genome Project , a 15-year international effort to locate all of the genes in the human genome and make them accessible for further biological study. Another goal of the project is to determine the complete sequence of the genome's 3 billion DNA nucleotide base pairs.

1992 Researchers at Lawrence Livermore National Laboratory and Lawrence Berkeley National Laboratory in California discover a gene present in 25 to 30 percent of the population that predisposes individuals to increased heart attack risk. Discovery of this marker for heart disease on chromosome 19 may make possible the development of a simple test to screen humans for susceptibility to heart disease.

England's Wellcome Trust joins the Human Genome Project.

1994 The Department of Energy launches its Microbial Genome Program.

Cracking the Code

1995 Craig Venter and colleagues at The Institute for Genomic Research in Maryland decode the first whole genome of a free-living single-cell organism, the influenza microbe, using the whole genome shotgun sequencing method.

1996 Ian Wilmut and other researchers at Scotland's Roslin Institute clone a sheep from the cell of an adult ewe. This non-sexually produced animal is named "Dolly." The complete genome of the *E. coli* bacteria is sequenced.

1998 The first complete genome sequence of a multicellular organism, the roundworm *C. elegans,* is published.

1999 The DOE Joint Genome Institute, a genome-sequencing center formed by Lawrence Livermore, Lawrence Berkeley, and Los Alamos national laboratories, dedicates its new production sequencing facility in Walnut Creek, California.

The complete genome of the *Drosophila* fruit fly is sequenced.

2000 Working drafts of the human genome are completed by the public International Human Genome Project and by Craig Venter's Celera Genomics, a private company.

2001 The draft human genome sequence is published in the journals *Nature* and *Science*. Twenty sequencing centers in six countries — China, France, Germany, England, Japan, and the United States — contribute to the project. Most of the sequencing is done by five major centers: the Wellcome Trust's Sanger Center in England, the DOE Joint Genome Institute in California, and three NIH-funded centers at Baylor College of Medicine in Texas, Washington University School of Medicine in Missouri, and the Whitehead Institute in Massachusetts.

2001–2002 DOE launches its "Genomes to Life" program.

As rapid, highly accurate sequencing techniques become readily available, the complete genomes of a wide variety of microbes and model organisms, including the mouse, pufferfish, malaria mosquito, and sea squirt, are sequenced and analyzed. Genome comparisons yield significant new insights into the causes and progress of disease, biological evolution, and the relationship between organisms and the environment.

2003 The finished human genome is published concurrent with the 50th Anniversary of the discovery of the double helix.

For more information:
DOE Joint Genome Institute: www.jgi.doe.gov
DOE Human Genome Project: www.ornl.gov/hgmis
NIH Human Genome Project: www.genome.gov
Genomes to Life Program: http://DOEGenomesToLife.org
Microbial Genome Program:

FOCUS

Continued from Insert, Page 1

ment and functioning, as well as how they interact with their environment. These insights will substantially further our understanding of biological systems and the mechanisms of evolution.

At this pivotal moment in the history of genomics, the U.S. Department of Energy, the DOE Joint Genome Institute (JGI), and the DOE national laboratories that formed the JGI in 1997 – Lawrence Livermore, Lawrence Berkeley, and Los Alamos – are well positioned to pursue the ultimate goal of modern biology: a fundamental, comprehensive, and systematic understanding of life.

The JGI's world-class genome sequencing capability, along with its growing work in comparative and functional genomics and the Department of Energy's microbial genomics and "Genomes to Life " programs, will play a central role in moving biology to this deeper level of understanding. Some of the key questions we will be trying to answer in the coming years are:

• What is the function of the 30,000 or so genes in the human genome?

• How are those genes regulated? In other words, what tells genes when and where to switch on and off?

• What is the role of the vast stretches of noncoding and repetitive DNA in the genome?

• What is the role of single nucleotide polymorphisms, or SNPs – changes in single nucleic-acid bases within the genome?

• What is the genetic basis for health and the pathology of human disease?

• What can DNA analysis tell us about risks of future illness?

One of the most effective ways to answer these questions is through the use of comparative genomics (see "Somewhere in the past, we're all related" elsewhere in this insert). Now that a complete human genome is available to us, we can learn a great deal about the location, function and regulation of both genes and noncoding segments of DNA by comparing the human genome with the genomes of a variety of "model organisms" – such as mice, fish, frogs and primates – that can be studied in the laboratory.

Today's different animals arose from common ancestors tens of millions of years ago. If segments of the genomes of two different organisms have been conserved – meaning the sequences are the same in both – over the millions of years since those organisms diverged, then we can expect that the DNA sequences within those segments probably encode important biological functions.

Comparative genomics studies in the future will rely heavily on the unique capabilities and strengths of the Joint Genome Institute. JGI is a national resource for sequencing, and it will continue to build upon that capability in the years to come; but it is also the substrate for important biological research.

As scientists at the DOE labs and the broader scientific community become users of JGI's unique resource of high-throughput sequence data, they will contribute both to its analysis and to strategies for converting this one-dimensional sequence data into three-dimensional biology. Our goal at JGI is to link genome sequencing to science; to take these double strands of data and use them to begin to answer many of the fundamental questions of biology, ecology and human health.

Eddy Rubin, Director of the DOE Joint Genome Institute, is also Director of the Genomics Division at Lawrence Berkeley National Laboratory.

Glossary

cDNA
Complementary DNA, a synthetic type of DNA generated from messenger RNA, or mRNA, the molecule in the cell that takes information from protein-coding DNA — the genes — to the protein-making machinery and instructs it to make a specific protein. By using mRNA as a template, scientists use enzymatic reactions to convert its information back into cDNA and then clone it, creating a collection of cDNAs, or a cDNA library. These libraries are important to scientists because they consist of clones of all protein-encoding DNA, or all of the genes, in the human genome.

centiMorgan
A unit of genetic distance. Generally, one centiMorgan equals about 1 million base pairs.

eukaryote
A single-celled or multicellular organism whose cells contain a distinct membrane-bound nucleus. If something is described as "eukaryotic," it means that it has cells with membrane-bound nuclei.

microarray
A device used in many types of large-scale genetic analysis. They can be used to study how large numbers of genes are expressed as messenger RNA in a particular tissue, and how a cell's regulatory networks control vast batteries of genes simultaneously. In microarray studies, a robot is used to precisely apply tiny droplets containing functional DNA to glass slides.

Researchers then attach fluorescent labels to complementary DNA (cDNA) from the tissue they are studying. The labeled cDNA binds to its matched DNA sequence at a specific location on the slide. The slides are put into a scanning microscope that can measure the brightness of each fluorescent dot. The brightness reveals how much of a specific cDNA fragment is present, an indicator of how active a gene is. Scientists use microarrays in many different ways. For example, microarrays can be used to look at which genes in cells are actively making products under a specific set of conditions, as well as to detect and/or examine differences in gene activity between healthy and diseased cells.

oligonucleotide
A short polymer of 10 to 70 nucleotides. A nucleotide is one of the structural components, or building blocks, of DNA and RNA. A nucleotide consists of a base chemical — either adenine (A), thymine (T), guanine (G) or cytosine (C) — plus a sugar-phosphate backbone. Oligonucleotides are often used as probes for detecting complementary DNA or RNA because they bind readily to their complements.

SNP
Single nucleotide polymorphism. SNPs - pronounced "snips" — are common, but minute, variations that occur in the human genome at a frequency of one in every 300 bases. That means 10 million positions out of the 3 billion base-pair human genome have common variations. These variations can be used to track inheritance in families and susceptibility to disease, so scientists are working hard to develop a catalogue of SNPs as a tool to use in their efforts to uncover the causes of common illness like diabetes or heart disease.

STS
Sequence tagged site, a short DNA segment that occurs only once in a genome and whose exact location and order of bases is known. Because each is unique, STSs are helpful in chromosome placement of mapping and sequencing data from many different laboratories. STSs serve as landmarks on the physical map of a genome.